

Cloning of cDNAs Encoding Two Related 100-kD Coated Vesicle Proteins (α -Adaptins)

Margaret S. Robinson

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Abstract. Coat proteins of \sim 100-kD (adaptins) are components of the adaptor complexes which link clathrin to receptors in coated vesicles. The α -adaptins, which are found exclusively in endocytic coated vesicles, separate into two bands on SDS gels, designated A and C (Robinson, M. S., 1987. *J. Cell Biol.* 104:887–895). Two distinct cDNAs (sequences 1 and 2) encoding the two α -adaptins were cloned from a mouse brain cDNA library. Southern blotting indicates that there is one copy of each of the two α -adapting genes, and that there are no additional closely related genes. Based on the size of the predicted protein products of the two genes (108 and 104 kD), the relative abundance of the two messages in brain and liver, and

the reactivity of a sequence 1 fusion protein with different antibodies, it was possible to conclude that sequence 1 codes for A and sequence 2 for C. The two protein sequences are strikingly homologous to each other (84% identical amino acids), the major difference being an additional stretch of 41 amino acids, rich in prolines and acidic residues, inserted into the COOH-terminal half of A. In situ hybridization carried out on mouse brain sections indicates that the same cell type may express both transcripts, but that their relative expressions vary. Antipeptide antibodies are now being raised to find out whether the proteins are localized in functionally distinct populations of endocytic coated vesicles.

MOST eukaryotic cells contain at least two types of clathrin-coated pits and vesicles. Coated vesicles that bud from the plasma membrane are involved in receptor-mediated endocytosis: they concentrate certain receptors and their bound ligands, allowing these proteins to be efficiently internalized. Endocytic coated vesicles also participate in the recycling of secretory granule membrane, and thus are particularly abundant in neuronal cells (Heuser and Reese, 1973). A second type of coated vesicle is associated with the Golgi apparatus and is thought to play a part in the targeting of newly synthesized proteins to lysosomes and secretory granules (Griffiths and Simons, 1986).

In addition to clathrin, receptors, and ligands, coated vesicles contain clathrin-associated proteins, which have at various times been called assembly proteins (Zaremba and Keen, 1983), 100K/50K complexes (Pearse and Robinson, 1984), and, most recently, adaptors (Pearse, 1988). The adaptors are believed to mediate the interaction between clathrin and the cytoplasmic tails of receptors (reviewed by Pearse and Crowther, 1987), and therefore to determine which membrane proteins are to be recruited into a coated vesicle for transfer to another membrane compartment of the cell and which are to remain behind.

Two types of adaptors have so far been identified, called HA-I and HA-II because of their relative elution from hydroxylapatite columns (Pearse and Robinson, 1984). Both have been shown to consist of a heterodimer of proteins of \sim 100 kD, recently named adaptins (Pearse, 1988), and two

associated smaller proteins, all apparently at a 1:1:1 molar ratio (Keen, 1987; Ahle et al., 1988). Each adaptor contains one copy of a β -adapting (β or β') and one copy of a specific adapting: an α -adapting in HA-II and a γ -adapting in HA-I (Ahle et al., 1988). The α -adaptins are the best characterized so far. Two major α -adapting bands can be seen on SDS gels of bovine brain HA-II adaptors, and these have been designated bands A and C (band B, which runs in between the two α -adapting bands, is the β -adapting). A and C have been shown to be closely related to each other on the basis of peptide mapping, NH₂-terminal sequencing, and monoclonal antibody cross-reactivity. However, while C appears to be expressed in all tissues, A has only been detected in brain (Robinson and Pearse, 1986; Robinson, 1987).

Antibodies against the adaptins in the HA-II adaptor provided the first clue that different adaptors might be found in different types of coated vesicles. Cells that were stained with polyclonal antisera which reacted with both α - and β -adaptins showed labeling of both plasma membrane and Golgi-associated coated vesicles, although different antisera gave somewhat different patterns (Robinson and Pearse, 1986). However, cells that were stained with monoclonal antibodies against the two α -adaptins (A and C), which did not cross-react with β , showed labeling exclusively of plasma membrane-associated coated vesicles (Robinson, 1987). Subsequently, Ahle et al. (1988) stained cells with monoclonal antibodies they had raised against β and γ . They showed that β is found in both Golgi and plasma membrane-associ-

ated coated vesicles, while γ occurs only in the Golgi subset of coated vesicles. These results indicate that adaptor HA-II, containing a heterodimer of β and either α_A or α_C , docks onto the plasma membrane, while HA-I, containing a heterodimer of β' and γ , is targeted to the Golgi apparatus. Such a mechanism could allow the cell to sort different sets of receptors into coated vesicles budding from different membrane compartments.

Many questions still remain to be answered. For instance, what are the precise functions of the different proteins making up the adaptor complexes? Are there additional types of coated vesicles in the cell, and if so, do they contain different adaptors? Why is there a special type of HA-II in brain, containing A instead of C? Are A and C distinct gene products, or are they alternatively spliced products of the same gene? Are they localized in different cells, or in different parts of the same cell?

As a first step toward addressing some of these questions, I have cloned and sequenced mouse brain cDNAs coding for adaptins α_A and α_C , the two related proteins found in the HA-II adaptor. cDNA clones encoding a β -adapitin have also been isolated; these are now being sequenced and characterized by P. Parham and co-workers and will be described in a subsequent paper.

Materials and Methods

Isolation and Characterization of Clones

Adaptins α_A and α_C were purified from bovine brain coated vesicles by Tris extraction followed by a combination of gel filtration, hydroxylapatite chromatography, preparative gel electrophoresis, and SDS/hydroxylapatite chromatography, as described (Robinson and Pearse, 1986). The pooled A- and C-containing fractions from the last column ($\sim 500 \mu\text{g}$ protein) were dialyzed against 20% ethanol containing 0.5 mM 2-mercaptoethanol, lyophilized, and dissolved in 500 μl 70% formic acid. Proteolysis was carried out with $\sim 5 \text{ mg}$ CNBr for 4 h at room temperature, after which the sample was lyophilized, dissolved in 100 μl 0.1% trifluoroacetic acid, and applied to a C18 Nucleosil microbore HPLC column. The peptides were eluted with a gradient of 0–50% CH_3CN in 0.1% trifluoroacetic acid, and the peak fractions were collected manually. Sequencing was performed with a gas phase sequencer (model 890B; Applied Biosystems, Inc., Foster City, CA) with on-line HPLC identification of the phenylthiohydantoin-derivitized amino acids.

Two separate mixed oligonucleotides were made for peptide 34 (see Table I):

5' AGCTGCTTCCAGCGCTGGAAGAAGTCCTGGCTTGCCAT 3'
 A T T A G

5' AGCTGCTTCCAGCGCTGGAAGAAGTCCTGGGATGCCAT 3'.
 A T T T G

Recombinant DNA techniques were generally those described by Maniatis et al. (1982). A bovine brain cDNA library constructed in $\lambda\text{gt}10$ was kindly donated by A. P. Jackson and P. Parham (Department of Structural Biology, Stanford University School of Medicine, Stanford, CA) (Jackson et al., 1987). Oligonucleotides were end labeled with ^{32}P and used to screen 5×10^5 plaques, washing the filters in $2 \times \text{SSC}$ at 42°C . Positive clones were plaque purified, digested with Eco RI to determine the size of the inserts, and Southern blotted onto nitrocellulose. Radioactive probes were made from the different inserts either by nick translation or by subcloning into M13 mp19 and synthesizing a radiolabeled second strand which could then be cut with a suitable enzyme and gel purified.

A $\lambda\text{gt}11$ mouse brain cDNA library, kindly provided by Y. Citri (Weizmann Institute of Science, Rehovot, Israel) (Martinez et al., 1987), was screened with the radiolabeled bovine brain inserts. Selected mouse brain inserts obtained from this screen were subcloned into a Bluescript plasmid vector (Stratagene Cloning Systems, San Diego, CA) for restriction mapping. Sequencing was carried out on consecutive and overlapping restriction

fragments subcloned into M13 mp18 or mp19, using the dideoxy chain termination method (Sanger et al., 1977). To sequence the entire coding region of both cDNAs in both directions, synthetic oligonucleotides were used as primers to fill in the gaps.

Northern and Southern Blots

For Northern blotting, poly A-containing RNA was purified from mouse brain and mouse liver by the lithium/urea method (Auffray and Rougeon, 1980), subjected to electrophoresis (15 $\mu\text{g}/\text{lane}$), and blotted onto Hybond paper. The blots were probed with M13 fragments and washed at 65°C in $0.1 \times \text{SSC}$. Poly A-minus RNA from each of the two tissues was used as a control. Labeling of the tissues was quantified by cutting out the lane of interest and measuring the radioactivity, and then subtracting the radioactivity present in a control lane as background. The mean ratios from three separate experiments were taken.

Southern blots were performed on genomic DNA purified from mouse liver, using nick-translated restriction fragments as probes. Low stringency labeling was carried out by incubating with the probe in 35% formamide at 42°C and washing in $2 \times \text{SSC}$ at 65°C , while the high stringency incubation was done in 50% formamide at 42°C , followed by a wash in $0.1 \times \text{SSC}$ at 65°C .

Production of Fusion Protein

The $\lambda\text{gt}11$ phage containing both sequence 1 and 2 inserts were streaked out onto a lawn of Y1090 host cells. Induction of fusion protein and labeling of the filters were carried out as described by Huynh et al. (1983). The primary antibodies used for screening were a mixture of the monoclonal antibody AC1-M11 (Robinson, 1987) and an affinity-purified rabbit polyclonal antiserum against total HA-II 100-kD proteins, prepared essentially as described (Robinson and Pearse, 1986). Phage that produced a signal were then used as lysogens to infect Y1089 cells (Huynh et al., 1983). Crude cell lysates were subjected to electrophoresis and blotted, and the blots were labeled with AC1-M11 (Robinson, 1987).

In Situ Hybridization

Freshly dissected adult mouse brains were frozen in dry ice, and 12- μm sections were cut using a cryostat. The sections were treated essentially as described by Goedert (1987), except that they were blocked with acetic anhydride after the pronase digestion step, and extracted with chloroform after serial dehydration in ethanol. Hybridization was carried out using single stranded ^{32}P -labeled M13 probes in a hybridization mix containing 50% formamide for 16–20 h at 42°C , and the slides were washed in $0.1 \times \text{SSC}$ at 60°C . The probes were prepared by subcloning Hind III/Eco RI fragments (bases 2,719–3,298 for gene 1 and bases 2,449–3,096 for gene 2) into both mp18 and mp19 and synthesizing a labeled second strand which was then cut with Eco RI (mp18) or Hind III (mp19). The mp19 probes were thus in the antisense direction, while the mp18 probes were in the sense direction and were used as controls. The slides were exposed on Fuji x-ray film for 3 d to make contact prints, and were then dipped in emulsion and exposed for 2–4 wk before being developed.

Results

Isolation of Two Distinct cDNAs

The approach that was taken for cloning α -adapitin cDNAs was to use oligonucleotides based on peptide sequences to screen a bovine brain cDNA library. The NH_2 -terminal sequences of both proteins have been determined and are known to be identical (Robinson, 1987). For internal sequencing, the α -adaptins were separated from β -adapitin by hydroxylapatite chromatography in the presence of SDS (Robinson and Pearse, 1986), digested with CNBr, and chromatographed on an HPLC column. Although there was some overlap in the elution profile, five of the peptide peaks gave useful sequence, and in some cases, both a major and a minor sequence could be determined. Table I lists the six sequences that were obtained.

Table I. Comparison of Bovine Peptides and Mouse Protein Sequences Deduced from the cDNAs

NH ₂ terminus/Peptide 10:	P	A	V	S	K	G	S?	G																		
Sequence 1:	P	A	V	S	K	G	D	G																		
Sequence 2:	P	A	V	S	K	G	D	G																		
Peptide 21:	G	L	A	L	H	N?	I	A	N	V	G	S	R	E												
Sequence 1:	M	C	L	A	L	H	C	I	A	N	V	G	S	R	E											
Sequence 2:	M	G	L	A	L	H	C	I	A	N	V	G	S	R	E											
Peptide 34:	D	S	V	K	Q	S	A	S	L	E	L	L	R	L	Y											
Sequence 1:	M	D	S	V	K	Q	S	A	A	L	C	L	L	R	L	Y										
Sequence 2:	M	D	S	V	K	Q	S	A	A	L	C	L	L	R	L	Y										
Peptide 26:	G	E	N	T	A	R	V	V	H	L	L	N														
Sequence 1:	M	G	E	W	T	A	R	V	V	H	L	L	N													
Sequence 2:	M	G	D	W	T	S	R	V	V	H	L	L	N													
Peptide 39:	W?	T	L	A	S	S	E	F	S	H	E	A	V	K	T	H	I	E	T	V	I	N	A	L	K	
Sequence 1:	M	C	T	L	A	S	S	E	F	S	H	E	A	V	K	T	H	I	D	T	V	I	N	A	L	K
Sequence 2:	M	C	T	L	A	S	S	E	F	S	H	E	A	V	K	T	H	I	E	T	V	I	N	A	L	K
Peptide 34:	A	S	Q	D	F	F	Q	R	W	K	Q	L	S	L	P	Q	Q	E	V	Q						
Sequence 1:	M	A	A	Q	D	F	F	Q	R	W	K	Q	L	S	L	P	L	Q	E	A	Q					
Sequence 2:	M	A	S	Q	D	F	F	Q	R	W	K	Q	L	S	N	P	Q	Q	E	V	Q					

Peptides 34 and 39 were both used to design synthetic oligonucleotides, but only peptide 34, which later was found to be nearer the COOH terminus of the protein, proved useful for library screening. A mixed oligonucleotide was made from this sequence, incorporating the first 12 amino acids plus a methionine codon at the 5' end. Nine positive clones were obtained, with inserts ranging in size from 0.5 to 1.2 kb. Interestingly, they appeared to fall into two categories. A probe made from the 0.5-kb insert hybridized well to itself but poorly to all the other inserts, while a probe made from one of the other inserts hybridized well to all but the 0.5-kb insert. When some of the inserts were sequenced, the reason became apparent: they coded for two different but closely related proteins, both of which contained sequences nearly identical to that of the peptide. Sequence 1 (Table I) differs from peptide 34 in that it has an alanine substituted for the serine at position 2 and an alanine for the valine at position 19; while sequence 2 has an asparagine substituted for the leucine at position 14. (The substitution seen at position 16 in sequence 1 does not occur in the bovine cDNA; Table I gives the mouse sequences 1 and 2.) For all but the ser/ala substitution, both amino acids could in fact be detected on the sequencer. The most likely explanation for this finding is that the cDNAs are the products of two genes, one of which codes for A and the other for C.

To obtain the complete sequences of the two proteins, a mouse brain cDNA library was used because it had longer inserts. The bovine brain cDNA inserts were used as probes, and 20–30 positive clones were isolated for each of the two transcripts. The sequences of the two mouse cDNAs are shown in Fig. 1, as well as their predicted protein products. A single screen yielded a clone containing the entire coding sequence of cDNA 2, while for cDNA 1, the 5' end of the coding region was obtained by rescreening the library with a Pst I fragment (bases 443–1,012) made from an insert obtained in the first screen. Restriction mapping was carried out on the longest inserts to look for possible polymorphisms

(e.g., alternative splicing). Although all of the inserts obtained for cDNA 2 produced similar restriction maps, 4 of the 15 cDNA 1 inserts that were mapped appeared to differ at their 3' or 5' ends. Sequencing these inserts indicated that they were the result of incomplete (rather than alternative) splicing: typical splice donor or acceptor sites were seen, and the putative introns contained in-frame stop codons.

Although the two cDNAs were cloned from mouse, all of the bovine peptide sequences can be easily identified (see Table I and Fig. 1). These include the NH₂-terminal sequence, beginning with the proline at amino acid position 2. The proline is immediately preceded by methionine in both cases, which probably serves as the initiator and is then cleaved off in the mature protein. The few slight discrepancies between the peptide and the cDNA sequences could be due either to species differences or to contaminating peptides coeluting from the HPLC. Moreover, some of the CNBr fragments of A and C may have eluted at slightly different positions, causing some of the peptide sequences (e.g., sequence 26) to correspond more closely to sequence 1 and others (e.g., sequence 39) to correspond more closely to sequence 2.

Sequence Analysis

Three protein databases were searched for proteins homologous to those encoded by the two cDNAs, but no significant matches appeared. However, as shown in Fig. 2, the two proteins are remarkably homologous to each other: 783 out of 934 amino acids are identical, and most of the substitutions are conservative ones. The homology is particularly striking towards the NH₂-terminus: only a single substitution (lys for arg) is seen for over 100 amino acids. Least homologous is the proline-rich region between amino acids 650 and 750 of sequence 1, where an additional stretch of 41 amino acids, including 11 prolines and 10 acidic residues, occurs in sequence 1 with no correlate in sequence 2.

Partial sequences have also been obtained for the 3' ends of the two bovine cDNAs. Out of 93 amino acids in bovine


```

1  MPAVSKGDMRGLAVFISDIRNCKSKAEAIKRIKELANIRSKFKGDKALDGYSK
1  MPAVSKGDMRGLAVFISDIRNCKSKAEAIKRIKELANIRSKFKGDKALDGYSK
56  KKYVCKLLFIFLLGHDIDFGHMEAVNLLSSNRKYTEKQIGYLFISVLVNSNSELIR
56  KKYVCKLLFIFLLGHDIDFGHMEAVNLLSSNRKYTEKQIGYLFISVLVNSNSELIR
111  LINNAIKNDLASRNPTFMCLALHCIANVGSREMGEAFADIPRILVAGDSMDSVK
111  LINNAIKNDLASRNPTFMGLALHCIANVGSREMAEAFAGEIPKILVAGDTMDSVK
166  QSAALCLLRLYKASPDLVPMGEWTAARVVHLLNDQHMGVVTAAVSLITCLCKKNPD
166  QSAALCLLRLYRTSPDLVPMGDWTSRNVHLLNDQHLGVVTAAATSLITTLAQKNPE
221  DFKTCISLAVSRLSRIVSASSTDLQDYTYFVPAPWLSVKLLRLLQCYPPPEDA
221  EFKTSVSLAVSRLSRIVTSASSTDLQDYTYFVPAPWLSVKLLRLLQCYPPP-DA
276  VKGRLVECLETVLNKAQEPKSKKVOHSNAKNAILFETISLIHYDSEPNLLVRA
276  VRGRLTECLETILNKAQEPKSKKVOHSNAKNAVLFEAISLIHYDSEPNLLVRA
331  CNQLGQFLQHRETNLRYLALESMTLASSEFSHEAVKTHIDTVINALKTERDVS
330  CNQLGQFLQHRETNLRYLALESMTLASSEFSHEAVKTHIE TVINALKTERDVS
386  RQRAADLLYAMCDRSNAKQIVSEMRLVLETADYATREEIVLKVAILAEKYAVDY
385  RQRAVDLLYAMCDRSNAQQIVASEMRLVLETADYSITREEIVLKVAILAEKYAVDY
441  WYVDTILNLIIRIAGDYVSEEVWYRVLQIVTNRDDVQGYAAKTVFEALQAPACHEN
440  WYVDTILNLIIRIAGDYVSEEVWYRVIQIVINRDDVQGYAAKTVFEALQAPACHEN
496  MVKVGGYILGFEFGNLIAGDPRSSPPVQFSLLSKPFHLCSVATRALLLSTYIKFIN
495  LVKVGGYILGFEFGNLIAGDPRSSPLIQFNLLSKPFHLCSVPTRALLLSTYIKFVIN
551  LFPETKATIQQVLRAGSQLRNADVLEQQRAVEYLRLSVASTDVLATVLEEMPPF
550  LFPVEKATIQQDLRSDSQLKNADVLEQQRAVEYLRRLS TVASTDILATVLEEMPPF
606  PERESSILAKLRK KKKGP GAASA LDDSR RDT S SN DI N GG V EPT P - - S TV ST P S P S
605  PERESSILAKLRK KKKGP S T V T D L E E T K R E R S I - D V N G G P E P V P A S T S A A S T P S P S
658  ADLLGLR A A P P - P A A P P A P V G G N L L V D V F S D G P T A Q P S L G P T P E E A F L S E L E P P A
659  ADLLGLG A V P P A P T G P P P S S G G L L V D V F S D S A S A V A P L A P G S E D N F - - - - -
712  P E S P M A L L A D P A P A A D P G P E D I G P P I P E A D E L L N K F V C K N S G V L F E N Q L L Q I G V K
- - - - - A R F V C K N N G V L F E N Q L L Q I G L K
767  SEFRQNLGRMYLFYGNKTSVQFQNFLLPTVVHPGDLQTQLAVQTKRVAQAQVDGGAQ
728  SEFRQNLGRMFI FYGNKTS T Q F L N F T P T L I C A D D L Q T N L N L Q T K P V D P T V D G G A Q
822  VQQVNLNIECLRDFLTPPLLSVRFRYGGTAQSLTLKLPVTTINKFFQPTEMA AQDFF
783  VQQVNLNIECLISDFTEAPVLLNIQFRYGGTFQNVSVKLPITL NKFFQPTEMAS QDFF
877  QRWKQLSLPLQEAQKIFKANHPMDAEVTKAKLLGFGSALLDNVDPNPE N F V G A G I
838  QRWKQLSNPQEQVQNI F K A K H P M D T E I T K A K I I G F G S A L L E E V D P N P A N F V G A G I
932  TQTRALQV GCLLRLEPNAAQAQMYRLTLRTSKEPVSRHLCELLAQQF
893  IHTKKTQIGCLLRLEPNLQAQMYRLTLRTSKDTVSRHLCELLSEQF

```

Figure 2. Comparison of the predicted protein sequences of the products of cDNA 1 (above) and cDNA 2 (below). The two proteins are 84% identical, with regions of greater and lesser homology. The N₂ termini are especially similar; while towards the COOH termini, an additional stretch of 41 amino acids is inserted into sequence 1.

be identified in the two proteins. Although it is not known which if any of these sites are used, studies have shown that at least some of the adaptins are phosphorylated *in vivo* (Keen and Black, 1986; Bar-Zvi et al., 1988).

Identity of the Two Encoded Polypeptides

Assuming that one of the two cDNAs codes for A and the other for C, which is which? Two lines of evidence indicate that cDNA sequence 1 codes for A and cDNA sequence 2 for C. First, the predicted molecular weight of the protein product of cDNA 1 is 107,605, while that of cDNA 2 is 104,017. Apparent molecular weights of A and C based on their mobility on SDS gels vary according to the gel system

used, but A consistently migrates more slowly than C, with molecular weights estimated at ~112,000 for A and 105,000 for C (Pearse and Robinson, 1984; Ahle et al., 1988).

Second, we have shown that C is expressed in many different tissues, while A has so far only been detected in brain (Robinson, 1987). To find out whether either of the two cloned cDNAs is brain specific, mRNA was purified from mouse brain and mouse liver, and equal amounts of each were subjected to electrophoresis and blotted. The blots were then probed at high stringency with labeled fragments from the 3' end of inserts containing sequences 1 and 2 (Fig. 3). The probe for sequence 1 labeled a single band in each tissue of ~4.5 kb, while the probe for sequence 2 labeled two bands in each tissue, of ~4 and 5.5 kb. Both probes labeled brain

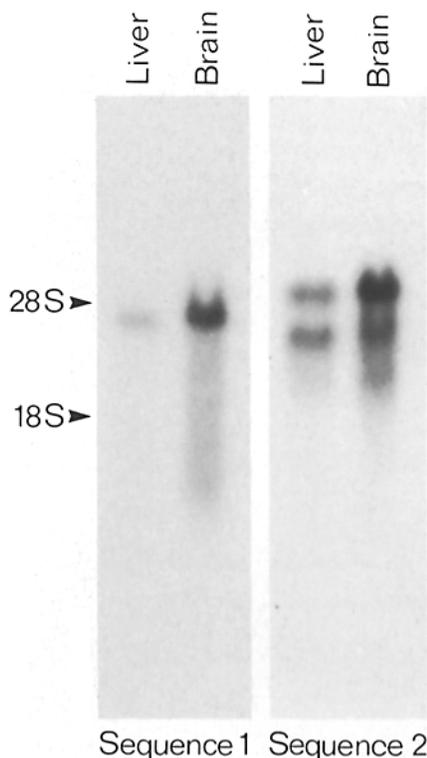


Figure 3. Northern blots carried out on mouse liver and mouse brain. Poly A-selected RNA was probed at high stringency with antisense DNA from the 3' ends of the two cDNA: bases 2,719–3,298 for sequence 1 and 2,449–3,096 for sequence 2. Positions of rRNA bands are indicated.

more strongly than liver, but when the labeling was quantified it was found that expression of sequence 2 in liver was 56% of that in brain, while expression of sequence 1 was only 9%. Western blotting of whole tissue homogenates indicates that C is more abundant in brain than in liver (Robinson, 1987), consistent with sequence 2 coding for C; but the finding that a small amount of sequence 1, presumably coding for A, is expressed in liver as well as in brain was unexpected. It may be that the amount of A produced in liver has so far escaped detection by Western blotting.

Another possible way of identifying the two cDNAs is by using monospecific antibodies to study their protein products. Since the cDNAs were cloned in the phage λ gt11, which can be used as an expression vector, there was a chance that some of the DNA might be inserted in such a way as to allow the expression of fusion protein. Accordingly, the phage were grown in a suitable host and the plaques tested for the presence of antigen. Two of the phage containing sequence 1 produced plaques that could be labeled both by an affinity-purified polyclonal antiserum and by monoclonal antibody AC1-M11, which recognizes both A and C (Robinson, 1987). Fig. 4 shows a gel and Western blot of *Escherichia coli* cells after a lysogenic infection with one of the two phage, labeled with AC1-M11. Although several labeled bands can be seen on the blot, indicative of proteolysis, the highest band runs with the mobility predicted for this particular construct. As expected, neither of the two C-specific monoclonal antibodies, which recognize different epitopes from each other (Robinson, 1987), labeled protein encoded by sequence 1.

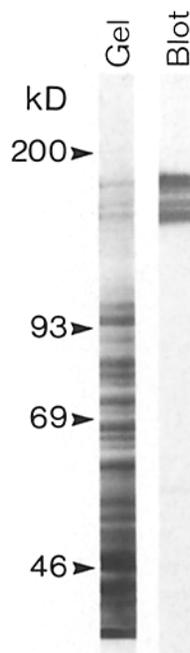


Figure 4. Western blot of bacteria expressing sequence 1. Cells were lysogenically infected with a λ -gt11 phage carrying part of cDNA sequence 1 fused to the β -galactosidase gene, and the blot was probed with a monoclonal antibody against the α -adaptins. Positions of molecular mass standards are indicated. The highest molecular mass band labeled with the antibody is \sim 180 kD, which is the predicted size for this particular fusion protein.

Although a positive result would be more meaningful than a negative result for this experiment, unfortunately none of the cells infected with the sequence 2-containing phage produced any detectable antigen. Nevertheless, the finding that protein encoded by sequence 1 reacts with antibodies against the α -adaptins provides additional and independent confirmation that the correct cDNA has been cloned.

Search for Other Genes

Are there any other genes that are homologous to those coding for A and C? When the bovine brain library was screened with oligonucleotides, only cDNA sequences 1 and 2 were isolated, but it is possible that the codon bias of the oligonucleotide prevented the detection of other, related cDNAs, or that such genes might not be expressed in brain. Thus, Southern blots of mouse genomic DNA, digested with five different restriction enzymes, were probed with 800–900-bp fragments derived from the 5' ends of sequences 1 and 2 under both high and low stringency conditions (Fig. 5).

At high stringency (Fig. 5, *a* and *c*), both of the probes produced fairly simple patterns, generally labeling one band strongly with weaker labeling of one or more additional bands, although only a single band is labeled in the two Eco RI tracks. These results indicate that there is a single copy of each of the two genes in the mouse genome, but that both genes contain restriction sites for most of the enzymes. In some cases, restriction sites can be found in the coding region (for instance, both cDNAs contain Pst I sites near the 5' end), while in other cases, the restriction sites presumably occur in introns.

At lower stringency (Fig. 5, *b* and *d*), each of the two probes labeled additional bands, but essentially all of these bands can be accounted for by cross-hybridization between the two genes. For instance, in the Hind III tracks, both of the probes labeled a heavy band and a light band at high stringency. At low stringency, an additional band appears in each of the Hind III tracks (Fig. 5, *b* and *d*, *asterisks*), and

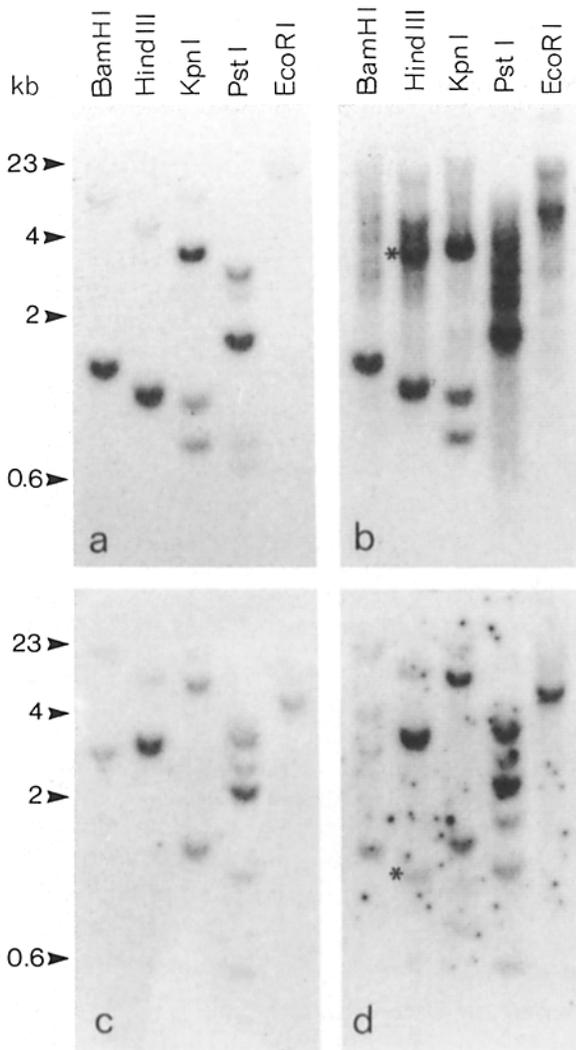


Figure 5. Southern blot of mouse genomic DNA probed at high and low stringencies. Mouse liver DNA was cut with the indicated restriction enzyme, and the blots were probed with fragments from the 5' ends of the two cDNAs: bases -205 to 708 for sequence 1 (*a* and *b*) and -150 to 683 for sequence 2 (*c* and *d*). High stringency labeling is shown in *a* and *c*; low stringency labeling is shown in *b* and *d*. Asterisks in *b* and *d* indicate bands appearing at low stringency in the Hind III track which can also be seen at high stringency with the other probe.

it exactly corresponds in size to the heavier band seen at high stringency with the other probe. Because no strong new bands appear, it seems unlikely that there are any other α -adaptin genes in mouse.

Localization of Transcripts

What function is served by having two different α -adaptins, one of which is mainly expressed in brain? One way of approaching this question would be to find out which cells in brain express sequence 1 and which express sequence 2, and whether the same cell can express both genes.

To localize mRNA for each of the two sequences, *in situ* hybridization was carried out on mouse brain sections, using antisense DNA as probes. Two pairs of consecutive serial sections are shown in Fig. 6, labeled with probes for se-

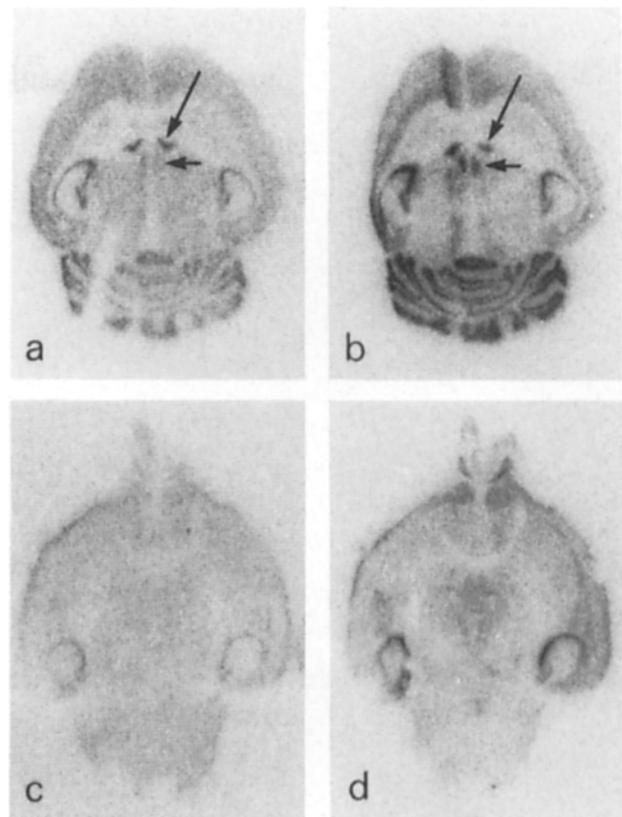


Figure 6. Contact prints of mouse brain sections hybridized *in situ* with probes for sequence 1 (*a* and *c*) and sequence 2 (*b* and *d*). Two pairs of consecutive serial sections from different parts of the brain are shown. The probes were the same as those used in Fig. 3. Arrows point to part of the dentate gyrus of the hippocampus (*long arrows*) and the habenular nucleus of the thalamus (*small arrows*) where differences in the intensity of the labeling can be seen.

quence 1 (Fig. 6, *a* and *c*) and sequence 2 (Fig. 6, *b* and *d*). Both probes labeled some parts of the brain more strongly than others, with the hippocampus, cerebral cortex, olfactory bulbs, and cerebellar grey matter particularly in evidence. These are all regions where neuronal cell bodies are concentrated, indicating that both genes are expressed in neurons. However, some differences can be seen in the two labeling patterns, one of which is indicated by the arrows. The long arrows point to part of the dentate gyrus of the hippocampus, while the short arrows point to the habenular nucleus of the thalamus. The probe for sequence 2 labeled both of these structures with about equal intensity, while the probe for sequence 1 gave relatively little labeling of the habenular nucleus. Control sections (not shown), which were incubated with sense DNA, showed only very faint uniform labeling.

The same slides were dipped in emulsion for resolution at the cellular level. Fig. 7 shows another region of the hippocampus from the same sections that appear in Fig. 6, *a* and *b*. Both probes strongly labeled the cells of the hippocampus, but the relative labeling of the two different cell types, the pyramidal and granule cells, differs in Fig. 7, *a* and *b*. The probe for sequence 2 (Fig. 7 *b*) labeled the pyramidal cells much more strongly than the granule cells, while the probe for sequence 1 (Fig. 7 *a*) labeled both types of cells more equally. Thus, the two genes appear to be expressed in the

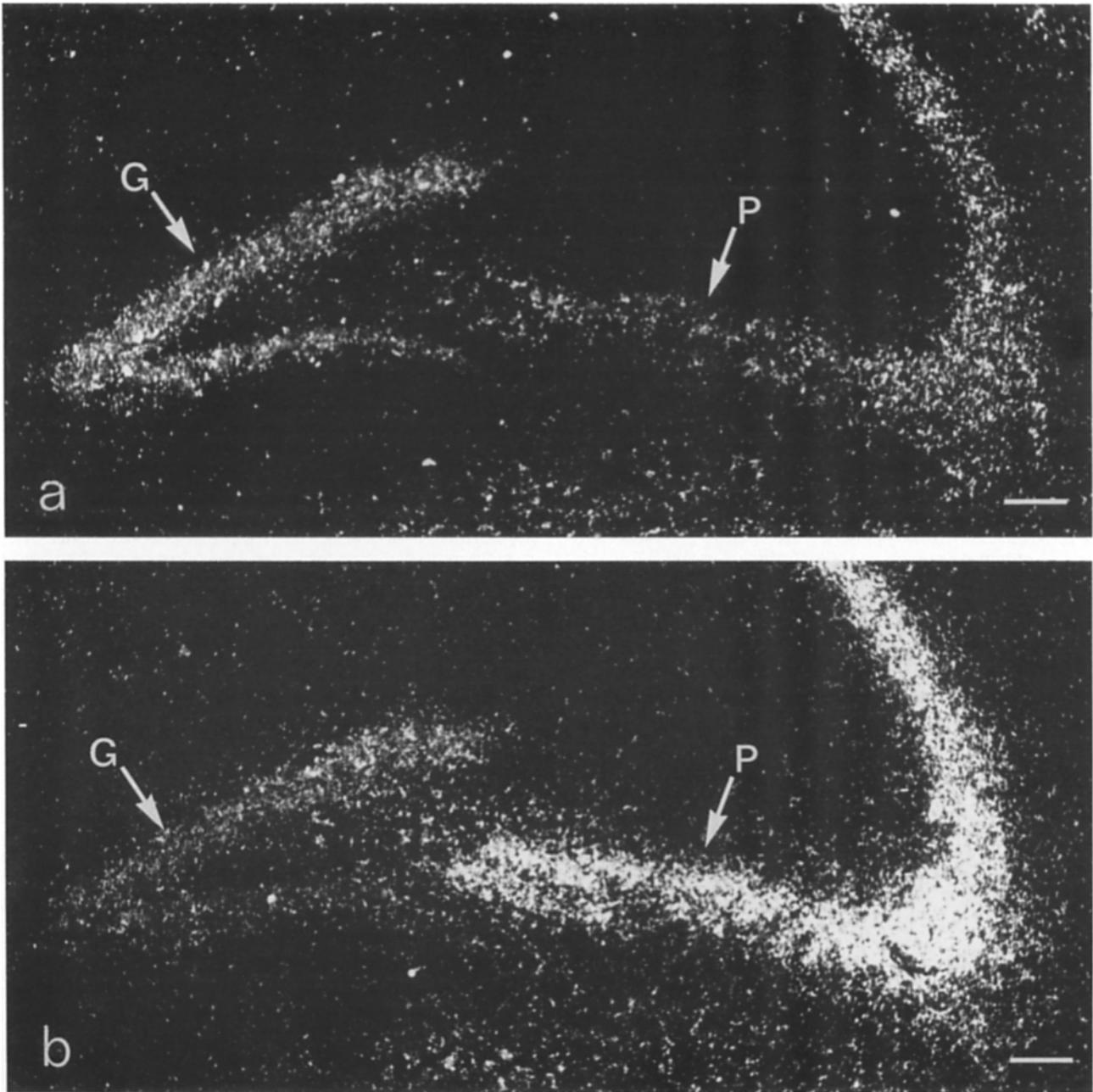


Figure 7. Microscopic view of the sections shown in Fig. 6, *a* and *b*. The granule cell layer (*G*) and pyramidal cell layer (*P*) of the hippocampus are indicated. Bars, 100 μm .

same populations of neuronal cells, but their relative expression differs in different cell types.

Discussion

Previous studies on the adaptins have made use of protein chemistry and monospecific antibodies to try to define the different types of proteins and their functions. In spite of the progress that has been made, many basic questions, such as how many different adaptins there are and how they are related to each other, have remained unanswered. By cloning and comparing the cDNAs coding for proteins, one can ad-

dress such questions directly; and in addition, tools then become available for learning more about protein function. Adaptins α_A and α_C are the first to be characterized in this way.

Before the cDNAs were cloned, there seemed to be a good chance that A and C would turn out to be alternatively spliced products of the same gene. The two proteins had been shown to be closely related, with identical NH₂-termini; and there were precedents for brain-specific splicing giving rise to a larger protein (Martinez et al., 1987; Jackson et al., 1987), consistent with the finding that A was only detectable in brain. However, the present study shows that the two pro-

teins are in fact encoded by two different genes. The two predicted protein sequences are highly homologous, sharing the same NH₂-termini as well as the five other peptides that were sequenced; and both the similarities and the differences that were observed in the two sequences are entirely consistent with biochemical and immunological results.

Three lines of evidence indicate that, of the two cDNAs that were cloned, sequence 1 codes for A and sequence 2 for C. First, sequence 1 encodes a larger polypeptide than sequence 2. Second, expression of sequence 1 in brain is about tenfold greater than in liver, while expression of sequence 2 is about twice that in liver. Finally, the protein product of sequence 1, expressed as a fusion protein, is recognized by antibodies that cross-react with A and C, but not by C-specific antibodies.

A computer search showed no significant homologies between the α -adaptins and any published protein sequences, nor has any homology been revealed so far between the sequences of the α and β -adaptins. One might expect γ -adaptin to be somewhat related to α , since it is thought to play an analogous role in the HA-I Golgi adaptor. However, it appears that A and C have no very close relatives, and thus are probably the only two α -adaptins. Southern blotting carried out under low stringency conditions, so that A and C are cross-hybridized to each other, did not reveal the presence of any new genes. Moreover, restriction mapping provided no evidence for alternative splicing within the coding region of the two genes. Thus, the observation that both A and C split into two bands on some gel systems (Ahle et al., 1988) is most likely due to posttranslational modifications such as phosphorylation. Interestingly, Northern blotting indicates that there are two species of mRNA for C. None of the clones that were isolated for C had a complete 3' end, as judged by the lack of a poly-A tail, but one such clone had a 3' untranslated region of ~ 1.5 kb; thus, this particular mRNA species must be over 4.5 kb. Since the two bands seen on Northern blots are ~ 4 and 5.5 kb, it is possible that there may be alternative splicing at the 3' untranslated end.

What, if any, are the functional differences between A and C? The only clue available at present is that A is primarily expressed in the central nervous system. However, although Western blotting has so far only detected A in brain and spinal cord (Robinson, 1987; and unpublished observations), Northern blots of liver mRNA show a small but reproducible amount of transcription of A as well as C. It is possible that only a subset of cells in liver express A: for instance, nerve cells, although the level of expression of A in liver seems to be too high to be attributable to nerve cells alone. In the case of brain, it has been possible to compare the overall distribution of the two messages by *in situ* hybridization. Both were shown to be most abundant in areas where neuronal cell bodies are concentrated. Moreover, one can conclude that the same neuronal cell types (e.g., the pyramidal cells of the hippocampus) express both genes, although the relative expression of the two genes varies according to cell type. Thus, although higher resolution studies are needed, it seems likely that the same cell can express both A and C. It is not clear how a cell would benefit from the coexpression of two closely related α -adaptins, but one possibility is that the proteins may be used to form two functionally distinct populations of endocytic coated vesicles. For instance, one α -adaptin may be used at the synapse and another in the rest of the cell, or

they may even be found in the same part of the cell but in different coated vesicles. Antibodies are now being made against synthetic peptides in order to compare the distribution of the two proteins in brain. It will also be important to find out whether A is translated in other tissues such as liver, and if so, whether it is present in the same coated vesicles as C.

It is known that different adaptors associate with different membrane compartments, but precisely how they do so is not clear. Pearse has shown that adaptors bind to the cytoplasmic tails of selected transmembrane proteins, such as the receptors for mannose-6-phosphate and low density lipoprotein (Pearse, 1985, 1988), but it seems unlikely that they are anchored by this interaction alone. The same proteins that are concentrated in coated pits on the plasma membrane are apparently excluded from coated pits during their biosynthesis when they pass through the Golgi apparatus (Griffiths et al., 1985). If adaptors were targeted to a particular membrane compartment solely through their binding to receptor tails, one would expect HA-II adaptors to be recruited to the Golgi apparatus as well as to the plasma membrane. Thus, it seems that there must be another docking mechanism to ensure that adaptors always go to the appropriate organelle. Clearly, more protein chemistry is needed to determine how the different proteins in the adaptor complex associate with each other, as well as with other components of the coated vesicle such as clathrin, receptors, and possible docking sites. The ability to express the cloned cDNAs in *E. coli*, where they can be experimentally manipulated, should be useful for such studies.

In addition, it should now be possible to learn more about adaptor function by expressing the α -adaptin cDNAs in tissue culture cells. With such an approach, one could try to create dominant mutants, or look at the role of phosphorylation, or find out whether the two α -adaptins are interchangeable by expressing A in cells that ordinarily only make C. Moreover, if γ -adaptin turns out to be related to the α -adaptins, as predicted, one could construct chimeric proteins to try to find out more about how they are targeted to the appropriate membrane compartment. For although we are beginning to understand something about how adaptors function in the sorting of other proteins, we still know nothing about how they themselves are sorted.

I thank Barbara Pearse for encouragement and advice; Fred Northrop for sequencing all the peptides; Tony Jackson, Peter Parham, and Yoav Citri for letting me use their cDNA libraries; Mike Runswick for help with the HPLC; Michel Goedert, Steve Hunt, Bill Wisden, and Maria Grazia Spilantini for teaching me how to do *in situ*; Tony Crowther for his computer expertise; Thierry Bogaert, Jon Glickman, Olof Sundin, and many others at the Laboratory of Molecular Biology for advice on techniques; and John Kilmartin for invaluable discussions.

Received for publication 19 August 1988 and in revised form 28 November 1988.

References

- Ahle, S., A. Mann, U. Eichelsberger, and E. Ungewickell. 1988. Structural relationships between clathrin assembly proteins from the Golgi and plasma membrane. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:919-929.
- Auffray, C., and F. Rougeon. 1980. Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* 107:303-314.
- Bar-Zvi, D., S. T. Mosley, and D. Branton. 1988. *In vivo* phosphorylation of clathrin-coated vesicle proteins from rat reticulocytes. *J. Biol. Chem.*

- 263:4408-4415.
- Goedert, M. 1987. Neuronal localization of amyloid beta protein precursor mRNA in normal human brain and in Alzheimer's disease. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3627-3632.
- Griffiths, G., S. Pfeiffer, K. Simons, and K. Matlin. 1985. Exit of newly synthesized membrane proteins from the *trans* cisternae of the Golgi complex to the plasma membrane. *J. Cell Biol.* 101:949-964.
- Griffiths, G., and K. Simons. 1986. The *trans* Golgi network: sorting at the exit site of the Golgi complex. *Science (Wash. DC)*. 234:438-443.
- Heuser, J. E., and T. S. Reese. 1973. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* 57:315-344.
- Huynh, T. V., R. A. Young, and R. W. Davis. 1983. Constructing and screening cDNA libraries in λ gt10 and λ gt11. In *DNA Cloning*. Vol. 1. D. Glover, editor. IRL Press Ltd, Oxford, UK. 49-78.
- Jackson, A. P., H.-F. Seow, N. Holmes, K. Drickamer, and P. Parham. 1987. Clathrin light chains contain brain-specific insertion sequences and a region of homology with intermediate filaments. *Nature (Lond.)*. 326:154-159.
- Keen, J. H. 1987. Clathrin assembly proteins: affinity purification and a model for coat assembly. *J. Cell Biol.* 105:1989-1998.
- Keen, J. H., and M. M. Black. 1986. The phosphorylation of coated membrane proteins in intact neurons. *J. Cell Biol.* 102:1325-1333.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
- Martinez, R., B. Mathey-Prevot, A. Bernardis, and D. Baltimore. 1987. Neuronal pp60^{c-src} contains a six-amino acid insertion relative to its non-neuronal counterpart. *Science (Wash. DC)*. 237:411-415.
- Pearse, B. M. F. 1985. Assembly of the mannose-6-phosphate receptor into reconstituted clathrin coats. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1257-2460.
- Pearse, B. M. F. 1988. Receptors compete for adaptors found in plasma membrane coated pits. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:3331-3336.
- Pearse, B. M. F., and R. A. Crowther. 1987. Structure and assembly of coated vesicles. *Annu. Rev. Biophys. Biophys. Chem.* 16:49-68.
- Pearse, B. M. F., and M. S. Robinson. 1984. Purification and properties of 100kd proteins from coated vesicles and their reconstitution with clathrin. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:1951-1957.
- Robinson, M. S. 1987. 100-kD coated vesicle proteins: molecular heterogeneity and intracellular distribution studied with monoclonal antibodies. *J. Cell Biol.* 104:887-895.
- Robinson, M. S., and B. M. F. Pearse. 1986. Immunofluorescent localization of 100K coated vesicle proteins. *J. Cell Biol.* 102:48-54.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463-5467.
- Zaremba, S., and J. H. Keen. 1983. Assembly polypeptides from coated vesicles mediate reassembly of unique clathrin coats. *J. Cell Biol.* 97:1339-1347.